Gene Hunters Close In On Elusive Prey

Researchers are deciphering the roots of diseases involving a complex mix of genes and environment—once thought impossible. The key: high technology and family sample collection on a grand scale

Two years ago, Noe Zamel, a geneticist at the University of Toronto, set out from Cape Town, South Africa, on an 8-day sea crossing through the violently stormy latitudes known as the "roaring forties." His destination: Tristan da Cunha, a tiny volcanic island in the middle of the South Atlantic Ocean—population 301. When he arrived, relieved, on dry land, he was greeted with a sign proclaiming, "Welcome to the loneliest island." Such an expedition may seem extreme for a research project, but the trip was well worth it. Tristan has an isolated, highly inbred population—all islanders are cousins—and one third of them suffer from asthma. Using DNA samples Zamel brought back, collaborator Lon Cardon at Sequana Therapeutics in La Jolla, California, is closing in on two genes that contribute to the islanders' asthma, and he's finding hints of other, weaker ones yet to be analyzed.

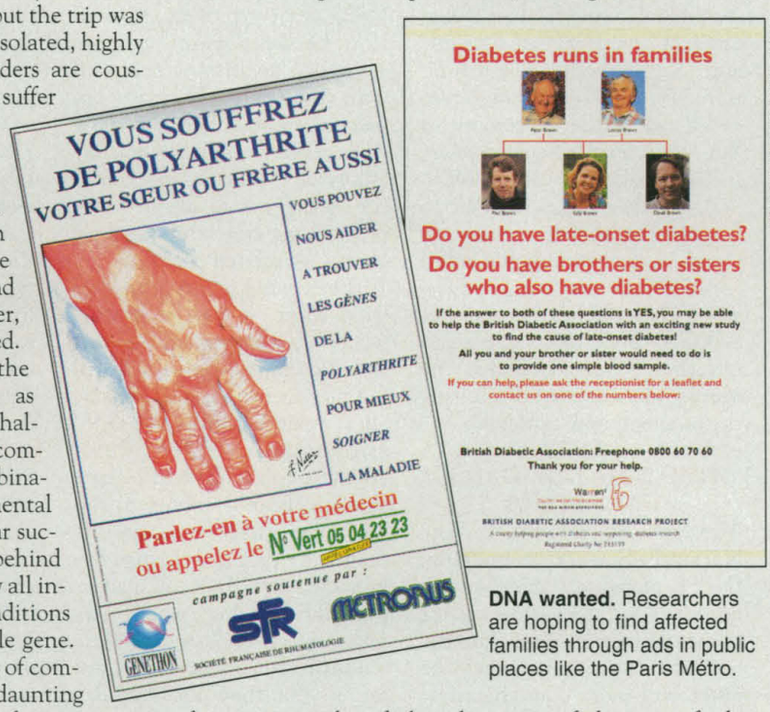
Zamel's trip symbolizes the difficulties researchers face as they take on the next big challenge in genetics research: complex diseases caused by a combination of genes and environmental factors. So far, the spectacular successes in finding the genes behind inherited diseases have nearly all involved relatively rare conditions caused by mutation in a single gene. Deciphering the genetic basis of complex diseases is a far more daunting task—so daunting that, just a few years ago, many researchers considered it almost hopeless. But, as participants gather for next week's Human Genome Organization meeting in Heidelberg, Germany, complex diseases are moving to the center stage of genome research. And as they do, researchers are forging new types of collaborations and—like Zamel—sparing no effort to secure the raw material for their studies: DNA from families with diseased members.

"This is where human genetics is at the moment," says genome researcher Peter Goodfellow of the University of Cambridge. "It is clear that, at least for some diseases, this will work." Indeed, in the past year, scientists have identified several genes that play a role in Alzheimer's disease, and teams on both sides of the Atlantic are making significant progress searching for others, including those behind asthma, juvenile and adult-onset diabetes, rheumatoid arthritis, and, more slowly, schizophrenia. "In the next year or 18 months there will be an explosion of knowledge," predicts asthma researcher William Cookson at the Wellcome Trust Centre for Human Genetics in Oxford.

The change in fortune has come about mostly through advances in genetic maps and technologies to exploit them, making it far easier to identify broad regions of the chromosomes where complex disease genes lurk. Although homing in on the culprits is still very difficult, greater knowledge of human genes, plus new technologies, will gradually resolve this. The big hurdle remains the need for DNA samples from large numbers of affected families.

This need stems from the difficulties in detecting complex disease genes, which increase susceptibility but do not inevitably lead to disease. What's more, each gene by itself usually has a small effect on risk, making it hard to spot—especially in outbred populations, where genetic variation between individuals is huge. And a gene that affects disease in one population may show little impact in another.

So it becomes a numbers game: The more families from diverse populations researchers can collect, the better their chances of finding the culprit genes. "Solid evidence [of a gene's location in the genome] will emerge from looking at different data sets. It's a statistical problem, a very hard one, unless you can get thousands of families," says geneticist Mark Lathrop, director of Wellcome's Oxford institute. "The clinical material is the real rate-limiting step," says Cookson. "The quality of the family material means the difference between success and failure."



DNA wanted. Researchers are hoping to find affected families through ads in public places like the Paris Métro.

Family fortunes

Researchers are scrambling to meet the growing demand for family samples with a broad palette of strategies. Many are teaming up in consortia that design a common protocol, then collect and pool their samples. Some groups are turning to major public appeals. Next month, a poster campaign to recruit families with rheumatoid arthritis will be launched in seven European countries, appearing everywhere from the Paris Métro to the phone booths of Madrid. And a U.S.-based consortium on asthma is seeking participants from specific ethnic groups through appeals in hospitals, at church fairs, on television, and even on the World Wide Web.

Researchers are also trying to find unusual, potentially more informative populations, such as the highly inbred people of Tristan da Cunha. Less genetic variation means a higher "signal-to-noise" ratio, so genes with weak effects stand out more against the background. Luckily, the two regions identified in Tristan inhabitants as most likely to contain asthma genes seem to have broader relevance, says Tim Harris, Sequana's vice president for research. "The Tristan population was the tool that opened the box," he says. "But we're fairly convinced that these genes are important for other populations, too."

The "right" populations were also key to finding genes that cause an early-onset form of Alzheimer's disease, which in some families is inherited much like a simple genetic disease. Pedigrees and DNA samples from

many afflicted families contributed to these landmark findings, but perhaps none so much as the family called "Hannah's heirs," after the book of that name by Daniel Pollen of the University of Massachusetts Medical Center in Worcester, one of the investigators involved in the research. Decades earlier, a pathologist in the family, realizing that the disease was inherited, had begun collecting the clinical records, diagnostic reports, and even tissue and blood samples which helped lead Peter St. George-Hyslop at the University of Toronto, plus a host of collaborators, to identify the gene last summer (*Science*, 30 June 1995, p. 1845).

Locating families to participate in the research isn't the only practical problem. "The clinical part [of complex disease gene searches] is very expensive and hard to find funding for," says Francis Collins, head of the U.S. National Center for Human Genome Research in Bethesda, Maryland. Collection projects have fared poorly under the U.S. peer-review system, which has tended to view them as "not very original," he says. "I think there's room to raise the consciousness of the funders ... [although] my impression is that they are now smiling a bit more on proposals to go out and collect." Some of the slack is being taken up by U.S.-based companies; for example, Sequana supports collection efforts of 20 academic institutions.

Collecting also takes a huge amount of experience and organization. Part of the effort is finding the specific types of families needed for mapping disease genes—typically those with at least two affected siblings (a small minority for most complex diseases) or large, multigenerational families with a higher-than-average incidence of the disease. "The major part of [gene mapping] work is designing the population you want to study, dealing with them as living human beings, and—very important—getting a correct diagnosis," says medical geneticist Albert de la Chapelle of the University of Helsinki in Finland. "Few people understand how work-intensive and difficult this is."

De la Chapelle knows firsthand because he and others have extensively studied Finland's genetically unusual population which, although far less inbred than Tristan's, originates from a fairly small number of people. That means many disease genes trace back to a single "founder"—so today's sufferers have identical copies not only of the disease gene but of the flanking regions. This, in turn, has helped researchers map genes for many single-gene diseases to regions of 100,000 bases or less—a high degree of precision that greatly simplifies zeroing in on the gene.

The hope that this shortcut will also work with complex disease genes is attracting gene hunters from around the world, as is Finland's superb infrastructure for genetic studies. One of those hunters is Collins, who, together with Jaakko Tuomilehto of Helsinki's National Public Health Institute and colleagues in the United States, is searching for genes involved in adult-onset diabetes. "Finland is an ideal place to do this work," says Collins. "They have a fantastic system of computerized medical records and a highly cooperative population that really believes in genetics research." That enabled the researchers to collect over 500 "sib-pair" families, plus a large body of clinical data, in 15 months—something of a speed record.

NOE ZAMEL

Worth the trip. Noe Zamel and colleagues are tracking down asthma-susceptibility genes from blood samples collected on the remote island of Tristan da Cunha.

Microsatellite mania

The technology that makes these searches possible is based on a monotonous-looking feature of the human genome: stretches of DNA called microsatellites, found all over the genome, where identical units of up to four nucleotides are repeated many times. Despite their dullness, microsatellites are ideal tools for mapping because, while their location in the genome is constant, they exist in alternate versions which vary in the exact number of repeats and are inherited as typical genetic traits. What's more, detecting them and working out the number of repeats at a specific site can be automated.

In 1990, with these tools in hand, Jean Weissenbach, Mark Lathrop, and a team of co-workers at Généthon—a major, charity-funded genome center in Evry, just outside Paris—set out to make a genetic map based on microsatellites. Using the industrial-scale setup at Généthon, within 2 years they located about 800 on the genome. Weissenbach continued after Lathrop moved to Oxford, and the latest map contains 5264 microsatellite markers—on average, one every 1.76 million bases. Two other teams—that of Ray White at the University of Utah in Salt Lake City and a consortium of over a dozen groups called the Cooperative Human Linkage Center, mapped several thousand more. With these tools, most simple disease genes can now be mapped in weeks—based on the idea that the closer a microsatellite lies to a disease gene, the more likely they are to be inherited together.

The same principle offered a way to map complex disease genes: the whole genome scan. In this approach, DNA from each family member is analyzed to see how many repeats are present at each of roughly 300 microsatellite loci distributed over the genome—a procedure called genotyping. With genotyping data from hundreds of people (needed for statistical significance), researchers can then ask questions such as whether sibling pairs with a particular disease share any microsatellites more often than expected by chance (50%)—a hint that a disease gene is located nearby. However, while the technology is straightforward, it is not small science: Mapping genes for a complex disease typically requires hundreds of thousands of genotypings, one reason it is largely the province of biotech companies and large-scale research centers (see box on p. 1354).

The first whole-genome scans in humans were carried out independently by the teams of John Todd and of Mark Lathrop and Cécile Julier, now all at Oxford's Wellcome Trust Centre, in 1993 to 1994. The researchers wanted to track down genes for juvenile-onset diabetes. Todd's preparations had begun in 1988, when he teamed up with the British Diabetic Association (BDA) and workers at the University of Birmingham to collect family material. In parallel with a Chicago-based group, the Human Biology Data Interchange, they established repositories to make the samples available to other researchers—the first such repository for a complex disease.

The results, published in *Nature* (8 September 1994, pp. 130 and 161), pointed to at least 10 regions of the genome that might contribute to diabetes. Two were already known from other approaches: the major histocompatibility complex (MHC) locus, which contains genes that regulate the immune response, and the region of the insulin gene itself. Finding them "reassured us that the method works," says Todd. But with just 300 families, he could only map down to regions still large enough to contain up to 500 genes—far too many to sift through to find the ones he is after.

That is just the point where most complex disease gene searches get stuck. But Todd has been able to home in further by identifying new microsatellite loci within the broadly mapped regions, then testing within families to see which ones are inherited most often by affected offspring and therefore must lie close to the culprit gene. This approach requires only one, not two, affected siblings—"a huge difference," says Todd. And it allowed him to map his way

Building a Home for Complex Disease Research

OXFORD, U.K.—The campus of the John Radcliffe Hospital in this quintessential university town has long been a focus for genetics research. It houses a formidable collection of facilities, including the University of Oxford's Institute of Molecular Medicine, a leading center on inherited blood diseases headed by Sir David Weatherall. And now there is a new kid on the block: the Wellcome Trust Centre for Human Genetics—one of the few academic institutes playing a leading role in research on the genetics of complex diseases.

The setting is crucial. It provides access to patients and clinicians, plus experience with large-scale collection of blood samples for analysis—a major bottleneck in research on complex disease genes (see main text). Precious expertise in designing large studies with patients also comes from the hospital's epidemiology and clinical trials groups.

KEITH BARNES/PHOTOGRAPHERS WORKSHOP

Multigene sleuths. Wellcome Centre Director Mark Lathrop *(center)* with colleagues.

The idea for the center came from a few researchers at the hospital who were already working in this growing field—including John Bell, John Todd, and William Cookson—plus Mark Lathrop, their frequent collaborator at Généthon in Paris. "We thought the time was ripe for the idea of funding a concerted activity around complex diseases," says Lathrop, now the center's director. So in 1992 they approached the London-based Wellcome Trust, the world's largest charitable funder of medical research. Wellcome liked the idea and committed $15 million in core funding for the first 5 years plus $30 million in grants. The center opened in June 1994 in renovated temporary lab space at the Nuffield Orthopaedic Centre.

One feature that sets the institute apart is its virtual production-line approach to tracking down susceptibility genes. Studies start out in core facilities where researchers collect patient samples, grow cells and extract DNA, then do the molecular analysis that roughly localizes the genes. A statistics group headed by Dan Weeks helps design the studies and interpret the results. "Extremely strong statistical genetics is crucial," says Lathrop. "Understanding what's a real effect and what isn't can be very tricky." Once a potential region is confirmed and narrowed down, the gene identification group helps with the often daunting job of going from a map location to specific candidate genes.

For molecular biologist Tony Monaco, this team approach "allowed me to do experiments I would never have dreamt of doing," he says. As a graduate student with Louis Kunkel at Harvard Medical School, Monaco worked on diseases caused by defects in single genes; he was the first to clone a gene (for Duchenne type muscular dystrophy) based on its position in the genome. Now he is tackling some of the most complex, least understood diseases: autism, language impairment, and other neurological disorders.

Down the hall from Monaco, immunogeneticist Adrian Hill works on susceptibility to infectious disease. Working with the U.K. Medical Research Council's unit in The Gambia, West Africa, he found that several genes of the immune system influence risk of malaria and tuberculosis, and is looking for others that play a role. He also participates in studies based in Tebbe, Uganda, on risk factors for HIV infection.

Lathrop hopes to round out the staff of about 150 with researchers from fields such as structural and developmental biology, to help understand the genes they eventually uncover. And by next year he should have the space to do it, in a new home for the center on the Radcliffe site that will put the search for complex disease genes on firm foundations.

–P.K.

right down to one of the unknown genes, which he is now identifying.

While collecting family samples for easy-to-diagnose diseases such as diabetes is difficult enough, it is far worse for psychiatric diseases such as schizophrenia and manic depression. "The biggest problem is that it's not entirely clear who has these diseases," says psychologist Donna Spiker of Stanford University, who coordinates collection for a project on autism. Diagnosis relies on analyzing the patient's behavior, which can be very subjective, and criteria change over time.

Despite the difficulties, many projects are now focusing on the genetic basis of psychiatric diseases. Newly published claims, for example, implicate a region on chromosome 6 in schizophrenia. That is turning lots of eyes to the next study expected to weigh in on the subject: the 14-country consortium coordinated by neurobiologist Jacques Mallet of the Hôpital de la Pitié Salpétrière in Paris. So far, scientists at Généthon have analyzed eight chromosomes from 1300 people from families with schizophrenic or manic-depressive members—the largest study so far, and the first to have rigorously standardized diagnostic criteria before collection, says Mallet.

Spotting susceptibility

Once researchers manage to zero in on a candidate gene, they face the next hurdle: proving it's the right one. Because a complex disease gene does not cause illness by itself, it can be present even in a healthy person. So proof again becomes a matter of statistical analysis, looking for differences in the gene in large numbers of affected and unaffected people. What's more, says Cookson, the difference between a normal and a susceptibility gene may be subtle—an issue he is struggling with. In 1994 he identified the only asthma gene yet known, which codes for a chain of the receptor for IgE, the "allergy antibody." But he has not yet found all the sequence differences that contribute to risk. "Many [susceptibility genes] won't be broken genes, but normal, common [gene variants]," he says.

But even before these genes are understood in all their subtleties, some researchers are putting the limited knowledge gained so far to work in large epidemiological studies aimed at disease prevention. One such project is DAISY—Diabetic AutoImmunity Study of the Young. For the past 18 months, pediatric endocrinologist Marian Rewers of the University of Colorado, Denver, has been screening children born in one city hospital for the known diabetes susceptibility genes at the MHC locus, which increase the risk of disease by as much as 25-fold, he says. The idea is to follow children who are at risk over many years and try to discover who actually gets the disease—and what environmental factors may have triggered it.

Getting this far is still a long way off for most complex diseases. "We're all on a learning curve," says Cookson. "At first, everyone was very unsophisticated. People didn't realize how many families you need, how difficult this is, and what a big, expensive undertaking." Now they know—but studies like DAISY offer hope that these long searches will, in the end, prove worthwhile.

–Patricia Kahn